A Machine Learning-Based Method for Automated Land Use Data Generation from Satellite Imagery

Kento Terada[1], Takahito Shimizu[2], Yuki Akiyama[3], Ryota Yamanotera[4]

[1] Tokyo City University, Setagaya-ku, Tokyo, Japan – g2581638@tcu.ac.jp
[2] Tokyo City University, Setagaya-ku, Tokyo, Japan – g2581624@tcu.ac.jp
[3] Tokyo City University, Setagaya-ku, Tokyo, Japan –akiyamay@tcu.ac.jp
[4] Tokyo City University, Setagaya-ku, Tokyo, Japan –yamapro1925@gmail.com

Keywords: Land use data, Automated classification, Satellite imagery, Machine learning, Remote sensing.

Abstract

Land use data, provided as open data by Japan's National Land Numerical Information (NLNI), has long served as a fundamental resource across various fields such as urban planning and disaster prevention. The dataset divides the entire country into 100-meter square mesh units and classifies each unit according to its land use purpose, enabling spatially detailed understanding of land use patterns. However, maintaining this dataset requires significant time and cost, as human operators visually interpret land use for each mesh by overlaying satellite imagery with geospatial data. As a result, it is difficult to update the dataset rapidly on a nationwide scale, leading to insufficient responsiveness to changes in land use. To address this issue, this study developed a method to automate land use classification using satellite imagery and improve the efficiency of land use data maintenance. Specifically, the method utilizes high-resolution and high-frequency observation data from Sentinel-2 and employs machine learning to automatically classify land use into four categories: Residential land, Other inhabitable land, Water bodies, and Forests and wastelands. As a result, the method enables high-accuracy generation of land use data and achieves significantly improved efficiency compared to conventional approaches. Furthermore, by integrating time-series satellite imagery, the method shows potential for flexibly responding to changes in land use.

1. Introduction

1.1 Background

Japan's National Land Numerical Information (NLNI) is an open platform provided by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT), offering integrated access to geospatial data across the country. The platform provides access to a wide range of geospatial information, including land use, topography, and transportation networks. According to the 2024 guidelines for the maintenance of NLNI, it is necessary to improve and update the data while addressing diverse user needs under constraints of limited budgets and human resources (MLIT, 2024). Therefore, the use of advanced technologies such as AI is encouraged to promote more efficient data maintenance.

Among the core datasets included in NLNI, the "Detailed Land Use Mesh Data" (hereinafter referred to as "land use data") plays a particularly important role. This dataset divides the entire country into 100-meter square mesh units and classifies each unit according to its land use purpose, enabling a spatially detailed understanding of land use patterns. Consequently, land use data is essential for supporting decision-making across a wide range of domains, including urban planning, infrastructure development, and disaster prevention. It also serves as a critical foundation for the realization of smart cities.

A concept closely related to land use is land cover. While land cover classification is based on the physical state of the Earth's surface, land use classification is defined by the intended purpose of land as determined by human socio-economic activities. Internationally, these two concepts are clearly distinguished due to differences in their classification objectives and targets, and each adopts an independent classification system.

For example, the Land Use and Land Cover (LULC) classification system developed by the United States Geological Survey (USGS) hierarchically categorizes land features, with Level 1 representing land cover and Level 2 representing land use (Anderson et al., 1976). Similarly, the Food and Agriculture Organization (FAO)'s Land Cover Classification System (LCCS) defines land cover as the observable physical state of the Earth's surface, whereas land use is classified based on anthropogenic factors, including legal and management perspectives (FAO, 2025).

As illustrated above, land use and land cover are defined as distinct classification systems at the international level. In particular, in Japan, where urban, rural, forested, and aquatic areas are often located in close proximity and intermingle, understanding land use based on human activities has proven to be more practical for administrative operations, planning, and disaster management. Consequently, land use classification has been given greater emphasis. As a result, numerous studies have been conducted utilizing land use data in the Japanese context.

For example, Koarai and Nakano (2017) analyzed the geographical characteristics and causal factors of tsunami damage during the Great East Japan Earthquake by integrating land use data with topographic, elevation, and tsunami-related datasets. Their study demonstrates that land use information is effective for disaster risk assessment.

In addition, Ohashi et al. (2024) combined land use data from 1976 to 2014 with population data from the national census to predict the spatial impacts of future population decline on land use across Japan, using machine learning techniques. This study highlights the potential of applying land use data to time-series analyses.

However, several challenges remain in the maintenance of land use data. Currently, approximately 34 million mesh units covering the entire country are manually classified by human operators through visual interpretation of satellite imagery. As a result, the process requires years of effort and incurs substantial costs, which in turn limit the frequency of updates. Moreover, although the high spatial resolution of 100-meter mesh units enhances the practical utility of the data for applications such as urban and disaster planning, it also contributes to increased labor and operational costs in the maintenance process.

Furthermore, Takayanagi (2017) pointed out that variations in classification accuracy among operators and the infrequent updates of the dataset hinder adequate responsiveness to changes in land use. Against this backdrop, the development of methods to automate and streamline land use classification is increasingly expected—both in Japan and internationally—as a fundamental technology to support urban and regional planning.

Therefore, there is a growing need to fundamentally reconsider the methods used to maintain land use data and to adopt more efficient and automated approaches. In particular, establishing a system that enables rapid and continuous data updates would facilitate the tracking of temporal changes in land use and development trends in cities and regions, thereby enhancing the utility of land use data as a foundational resource for policy-making. Such approaches are expected to become increasingly practical in fields such as smart city development, disaster response, and urban reconstruction.

1.2 Previous Studies

In the field of remote sensing, techniques for the automatic classification of land cover and land use using satellite imagery have advanced significantly in recent years. In particular, the introduction of machine learning and deep learning has enabled high-accuracy classification of large-scale spatial data.

Li et al. (2018) conducted a systematic review of deep learning applications to remote sensing imagery and reported that deep learning models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs) outperform traditional hand-crafted feature-based methods in major tasks such as image classification, segmentation, and object detection. Zhu et al. (2017) also demonstrated that, compared to conventional machine learning techniques such as support vector machines (SVMs) and Random Forests, deep learning enables automatic and high-dimensional feature extraction from large-scale datasets, achieving superior performance. On the other hand, they also highlighted several challenges, including the need for large amounts of high-quality labeled data and the impact of class imbalance on learning performance.

Gong et al. (2020) demonstrated that, in scene classification using public remote sensing datasets, not only model selection but also factors such as image resolution and differences in target regions and structures have a significant impact on classification accuracy. Talukdar et al. (2020) compared machine learning algorithms including Random Forest, support vector machines (SVMs), and artificial neural networks (ANNs), and reported that Random Forest offers a favorable balance of classification accuracy, generalizability, and computational efficiency. Similarly, Ghayour et al. (2021) showed that Random Forest outperformed other methods in land cover classification using Sentinel-2 imagery, highlighting the effectiveness of combining it with spectral information.

On the other hand, the majority of these studies have focused on land cover classification, where the target classes typically consist of physical attributes such as forests, water bodies, and bare land. In addition, the classification units are generally either pixel-based or object-based, and the spatial scope tends to be limited to regional scales. Campos-Taberner et al. (2020) attempted to improve the classification accuracy of agricultural land use by applying a long short-term memory (LSTM) model to time-series Sentinel-2 imagery; however, the study was restricted to agricultural regions and employed a custom-defined classification scheme.

Although relatively few in number, some studies have focused on land use classification. For example, Zhao et al. (2019) integrated satellite imagery with social sensing data such as point-of-interest (POI) information and human mobility data to estimate land use categories—such as residential, commercial, and industrial areas—within urban environments. However, such approaches tend to heavily rely on human activity data, and their classification schemes do not necessarily align with administrative land use categories.

In Japan, Ochi (2009) applied an object-based approach to high-resolution satellite imagery and demonstrated improved delineation of land cover boundaries and enhanced classification accuracy in urban areas. This suggests that the choice of classification unit has a significant impact on the final classification results.

In summary, existing studies have largely focused on the classification of physical attributes or analyses conducted within limited geographic areas. As a result, technologies for automatically classifying and generating land use data at the national level, based on mesh units and aligned with administrative classification schemes, have yet to be fully established.

1.3 Objectives

The objective of this study is to establish a method for efficiently and frequently maintaining land use data across Japan by utilizing satellite imagery and machine learning. In particular, the study aims to develop a technique that enables the automatic classification of land use at a nationwide scale, based on 100-meter mesh units and aligned with the institutional classification scheme of NLNI provided by the Ministry of Land, Infrastructure, Transport and Tourism.

Based on the land use data from NLNI, this study consolidates land use categories into four classes: residential land, other inhabitable land, water bodies, and forests and wastelands. This classification scheme is designed to balance practical efficiency and classification accuracy by prioritizing the identification of major land use types, taking into account the geographical characteristics of Japan, where forests and water bodies occupy the majority of the national territory.

To evaluate the effectiveness of the proposed method, land use data provided by NLNI are used as reference data representing past interpretations, and the classification performance is assessed by comparing the predicted results with these reference data for two regions—Akita and Nara Prefectures. This evaluation enables the assessment of classification accuracy and generalization performance, including regional variation.

This paper is organized as follows: Section 2 describes the data and methodology used in this study; Section 3 presents the classification results and accuracy evaluation; Section 4 discusses the findings and identifies remaining issues; and Section 5 provides the conclusion and future perspectives.

2. Methodology

2.1 Study Areas

This study focused on two regions in Japan. The model was trained and validated with data from Akita Prefecture, and extrapolation testing was conducted using data from Nara Prefecture. Their locations are shown in Figures 1 and 2.

Akita Prefecture is located in one of the colder regions of Japan and is characterized by buildings with varied elevation profiles designed to accommodate heavy snowfall. In contrast, Nara Prefecture is situated in a warmer region and is characterized by sturdy, flat-roofed buildings designed to withstand typhoons. Such climatic differences are believed to influence building structures and land use patterns, and validation under varying climate conditions is considered effective for enhancing the generalizability of the proposed method.

Another notable feature of both regions is that they exhibit clearly defined land use patterns separating urban and rural areas. Therefore, these regions were selected as study areas to evaluate the general applicability of the proposed method.

2.2 Source Data

In this study, Sentinel-2 satellite imagery was used as the source of remote sensing data. Sentinel-2 is an Earth observation satellite operated by the European Space Agency (ESA), offering high spatial resolution of up to 10 meters, high revisit frequency of approximately every five days, and 13 spectral bands. The data

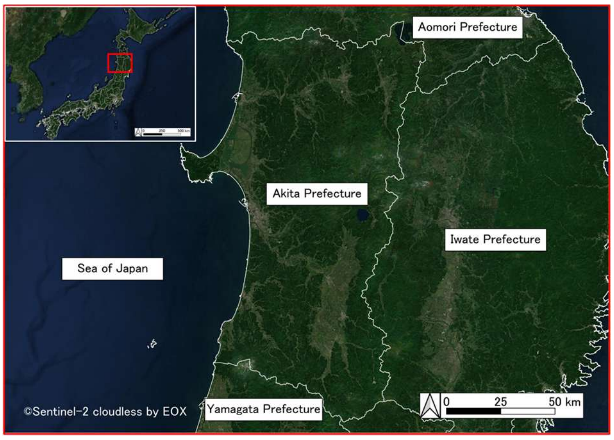

Figure 1. Study area: Akita Prefecture

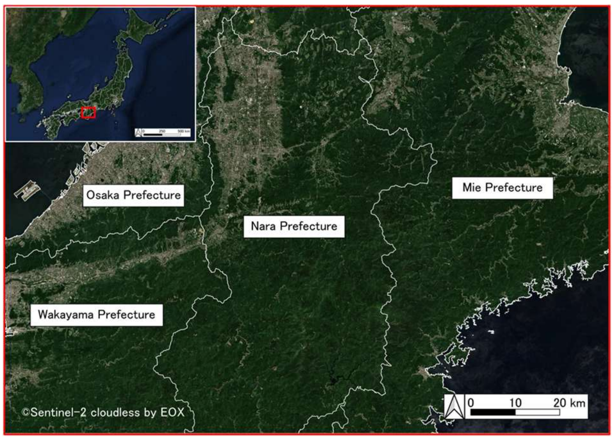

Figure 2. Study area: Nara Prefecture

are available free of charge (ESA, 2015). Owing to these characteristics, Sentinel-2 data have been widely used in various fields such as urban planning, agriculture, disaster management, and mining. In this study, they were adopted as suitable remote sensing data for land use classification.

For both model training and validation, interpreted land use data provided by the NLNI were used as training and evaluation data, respectively. In Akita Prefecture, the satellite imagery was aligned with the interpretation period of the land use data, which was November 2020. The scene with the least cloud coverage from within a one-month window before and after this period was selected using the STAC API provided by Amazon Web Services (AWS) (ESA, 2015).

In contrast, for Nara Prefecture, priority was given to data quality, and a 2023 annual mosaic image was used. For the sake of implementation efficiency and visual clarity, the Sentinel-2 Global Mosaic provided by the European Commission's Copernicus Programme was adopted (European Commission, 2023). The classification scheme of the NLNI land use data is described in detail in Section 2.4.

By using satellite imagery acquired from different regions and time periods, this study enables the examination of whether the model can effectively distinguish variations in land use within the imagery. Furthermore, it allows for the evaluation of the model's potential to accommodate temporal changes in land use.

2.3 Data Preprocessing

In this study, the acquired imagery underwent a series of preprocessing and feature extraction steps for land use classification. First, spatial clipping was performed on the satellite imagery based on the spatial extent of the land use data of the target regions in order to remove areas outside the scope of analysis. Next, multiple spectral bands and indices known to be effective for land cover identification were selected and used as input features for the classification model. The features used in this study are listed in Table 1.

Feature	Resolution (m)	Description
Blue (Band2)	10	Visible blue band (490 nm)
Green (Band3)	10	Visible green band (560 nm)
Red (Band4)	10	Visible red band (665 nm)
NIR (Band8)	10	Near-infrared band (842 nm)
SWIR (Band11)	20	Shortwave infrared band (1610 nm)
NDVI	10	Index that indicates the density and vigor of green vegetation
MNDWI	20	An index that highlights surface water (including snow) and reflects the moisture content in vegetation
NDSI	20	An index that detects the distribution of bare ground and impervious surfaces, such as sand and concrete

Table 1. Features of this study and their descriptions

As shown in Table 1, Blue, Green, and Red correspond to the visible light bands, specifically Band 2, Band 3, and Band 4, respectively. NIR refers to near-infrared and corresponds to Band 8, while SWIR denotes short-wave infrared and corresponds to Band 11. All bands were acquired at their highest available spatial resolution. In addition, three indices were derived from these bands and included as additional features.

First, the NDVI (Normalized Difference Vegetation Index) is an index used to evaluate vegetation presence and physiological activity, based on the spectral characteristics whereby vegetation exhibits high reflectance in the near-infrared region and absorbs strongly in the red region (Tucker, 1979).

Next, the MNDWI (Modified Normalized Difference Water Index) is an index used to enhance the extraction of water bodies, based on the spectral characteristics whereby water exhibits high reflectance in the green band and is strongly absorbed in the shortwave infrared region (Xu, 2006).

Finally, the NDSI (Normalized Difference Soil Index) is an index used to detect non-vegetated areas such as sand and concrete, based on the spectral characteristics whereby dry minerals and artificial structures exhibit high reflectance in the shortwave infrared region and vegetation is strongly absorbed in the near-infrared region (Chen et al., 2021). The NDSI helps emphasize non-vegetated areas and artificial structures, thereby improving the separability from water bodies and vegetation.

These indices were computed using the formulas shown below.

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \tag{1}$$

$$MNDWI = \frac{(Green - SWIR)}{(Green + SWIR)} \tag{2}$$

$$NDSI = \frac{(SWIR - NIR)}{(SWIR + NIR)} \tag{3}$$

2.4 Preparation of Training Data

In this study, the land use data from the NLNI were used as previously interpreted land use classification results, and training data for model development and evaluation were prepared accordingly.

The spectral features extracted in Section 2.3 had varying spatial resolutions; therefore, all features were resampled to 20 meters. Then, pixel values were averaged within each 100-meter mesh to obtain representative values for each mesh unit. To ensure spatial consistency between features and labels, each pixel was assigned to a 100-meter mesh based on its center coordinate, and the land use classification label assigned to that mesh was used.

Based on the design policy described in Section 1.3, the classification labels were restructured by aggregating the land use categories from the NLNI data into four classes: residential land, other inhabitable land, water bodies, and forests and wastelands. The correspondence between the original categories and the restructured classes is shown in Table 2. This classification scheme introduces a two-level hierarchy: the first level divides areas into inhabitable and uninhabitable zones, and the second level further splits each category into two sub-classes, resulting in a total of four classes.

In this study, inhabitable land is defined as land that has already been developed or could potentially be converted into residential areas in the future, such as agricultural land, roads, and railway tracks. In contrast, uninhabitable land refers to natural areas such as water bodies and forests, which are generally unsuitable for construction or permanent human settlement (Fujimoto, 2013). Inhabitable land is further divided into “residential land,” where buildings are present, and “other inhabitable land.” Uninhabitable land is subdivided into “water bodies” and “forests and wastelands.”

Through the above procedures, a training dataset was constructed in which each 100-meter mesh unit has a unified feature vector and corresponding label. This dataset serves as the foundation for training, validation, and extrapolation evaluation of the classification model described in the following sections.

2.5 Machine Learning Model Construction

Using the training dataset constructed in Section 2.4, a machine learning model was developed to automate land use classification across the entire Akita Prefecture. The primary configuration settings used for model training are summarized in Table 3.

Stage 1	Stage 2	Classification categories
Inhabitable land	Residential land	-
	Other inhabitable land	Paddy fields, Other agricultural land, Roads, Railways, Other land, Golf courses
Uninhabitable land	Water bodies	River and lake bodies, Coastal areas, Marine waters
	Forests and wastelands	-

Table 2. Land use classification mapping table

Item	Setting
Target area	Entire Akita Prefecture
Classification model	Random Forest
Number of training mesh units	Residential land: 38,831
	Other inhabitable land: 195,778
	Water bodies: 147,609
	Forests and wastelands: 1,115,886
	Total: 1,498,104
Training method	Five-fold cross-validation
Hyperparameters	n_estimators = 200
	max_depth = 10
	min_samples_split = 2

Table 3. Main configuration settings used for model training

https://doi.org/10.5194/isprs-archives-XLVIII-4-W16-2025-121-2025

In this study, the Random Forest algorithm (Breiman, 2001) was adopted as the base model. Random Forest is an ensemble learning method composed of multiple decision trees, and it demonstrates robust classification performance for high-dimensional and non-linear features. It is also resistant to class imbalance, making it well-suited for complex and structured classification tasks such as land use classification.

In recent years, studies have reported using Transformer-based deep learning models for automatic land use and land cover classification, showing high performance (Khan et al., 2024). However, such models require large training datasets, significant computational resources, and offer limited interpretability. Therefore, this study adopts Random Forest as the machine learning model, aiming to balance classification accuracy and computational efficiency.

In addition, LightGBM (Ke et al., 2017) was also tested for comparison. However, for the 100-meter mesh-based classification problem addressed in this study, LightGBM produced slightly less stable results compared to Random Forest. This outcome is likely due to LightGBM's tendency to be overly sensitive to subtle scale variations in the optimization of predictive performance, whereas Random Forest, with its simple majority-voting mechanism among decision trees, maintains more robust classification performance even in the presence of overlapping class distributions. Therefore, in this study, the Random Forest was ultimately used as the main model.

During model construction, approximately 1.5 million 100-meter mesh units covering the entire Akita Prefecture were used. Given the class imbalance present in the mesh label distribution, five-fold cross-validation was adopted. In this approach, the dataset was split into five subsets while maintaining class proportions within each fold; in each iteration, 80% of the data was used for training and the remaining 20% for validation. This method helps to prevent overfitting while ensuring generalization performance. Hyperparameters were set based on values commonly used in previous studies, such as Talukdar et al. (2020), and were found to yield satisfactory performance in this study as well.

Instead of performing a single multi-class classification for the four land use categories, this study employed a hierarchical two-stage binary classification approach. In the first stage, mesh units were classified as either inhabitable land or uninhabitable land. In the second stage, inhabitable land was further classified into residential land and other inhabitable land, while uninhabitable land was classified into water bodies and forests and wastelands. Each binary classifier outputs a probability between 0 and 1 for each class, with the higher value determining the predicted label for each mesh unit. This approach improves class separability by sequentially dividing land categories with distinct spectral characteristics and helps reduce the accumulation of classification errors. Furthermore, the hierarchical structure offers practical advantages by enabling a more interpretable and logically structured classification process.

2.6 Prediction and Evaluation

Using the machine learning model constructed in Section 2.5, land use classification was performed at the 100-meter mesh level for areas covering the entirety of Akita and Nara Prefectures and their surrounding regions. In this study, since marine waters are also included in the classification, the entire image, including surrounding areas, was used as the prediction target, rather than being separated by the administrative area of each region. The resulting predictions were quantitatively evaluated by comparing them with previously interpreted land use classification data provided by the National Land NLNI. For Akita Prefecture, internal validation was conducted using five-fold cross-validation, while for Nara Prefecture, generalization performance was assessed through extrapolation testing of the trained model.

In this study, the evaluation metrics used were overall accuracy, precision, recall, and F1 score. Overall accuracy represents the proportion of correctly classified mesh units out of the total. Precision refers to the proportion of correctly predicted instances among all instances predicted as a given class. Recall measures the proportion of correctly predicted instances among all instances that actually belong to the target class. The F1 score is the harmonic mean of precision and recall, serving as a metric that accounts for the balance between them. It should be noted that overall accuracy is defined as a single value across all classes in multi-class classification, and cannot be meaningfully assigned on a per-class basis. Therefore, only the overall value is reported.

3. Results

3.1 Classification Results in Akita Prefecture

Figure 3 presents the land use classification results for the four categories across Akita Prefecture and its surrounding regions. The evaluation metrics and classification accuracy for each class are summarized in Table 4. To enable a more detailed comparison of classification performance, a close-up view focusing on the western area of Akita City was generated. The corresponding classification results are shown in Figure 4, and the reference labels are shown in Figure 5. A visual comparison of these figures confirms a high degree of agreement between the predicted results and the actual interpretation labels.

3.2 Classification Results in Nara Prefecture

Figure 6 presents the land use classification results for the four categories across Nara Prefecture and its surrounding regions. The evaluation metrics and classification accuracy for each class are summarized in Table 5. To enable a more detailed comparison of classification performance, a close-up view focusing on the central area of Nara City was generated. The corresponding classification results are shown in Figure 7, and the reference labels are shown in Figure 8. Even in the extrapolation evaluation, a visual comparison confirms a high degree of agreement between the predicted results and the actual interpretation labels.

4. Discussion

4.1 Evaluation of Classification Performance

As shown in Figures 3 and 6, despite targeting large land areas, inhabitable and uninhabitable zones are clearly distinguished, indicating that the model successfully achieves detailed land use classification.

According to Table 4, the classification performance in Akita Prefecture shows a high overall accuracy of 88.2%. Focusing on each class, the indicators for uninhabitable areas all exceeded 80%, demonstrating strong performance. However, the precision for inhabitable areas did not reach 80%. In particular, the precision for residential land was extremely low at 32%. This may be attributed to the fact that uninhabitable areas occupy the majority of the target region, and among inhabitable areas, residential land had a relatively small number of samples available for training compared to the other classes.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W16-2025-121-2025 |

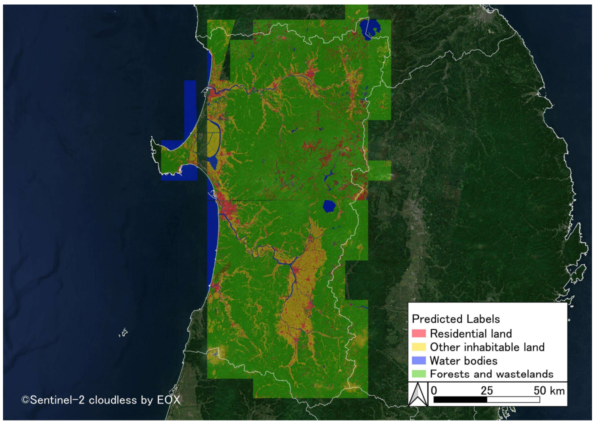

Figure 3. Land use prediction in Akita Prefecture

Akita Prefecture	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Residential land	-	32	86	46
Other inhabitable land	-	70	83	76
Water bodies	-	88	88	88
Forests and wastelands	-	98	89	94
Overall	88.2	72.0	86.5	76.0

Table 4. Evaluation metrics in Akita Prefecture

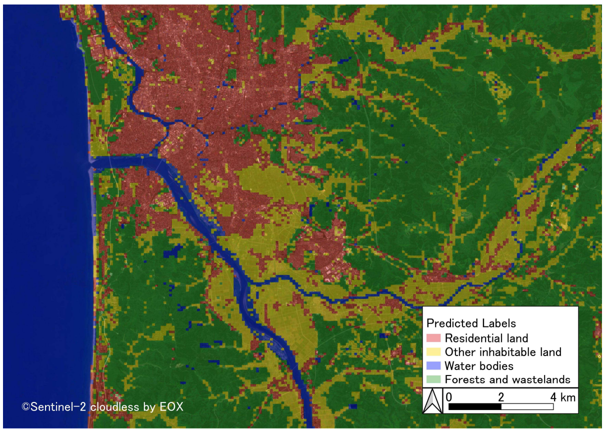

Figure 4. Predicted results in Western Akita City

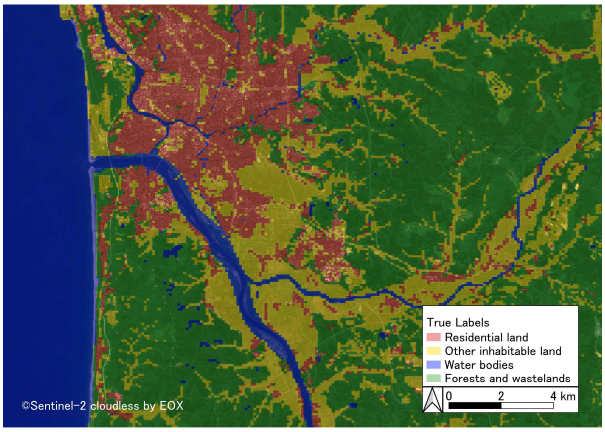

Figure 5. Ground truth labels in Western Akita City

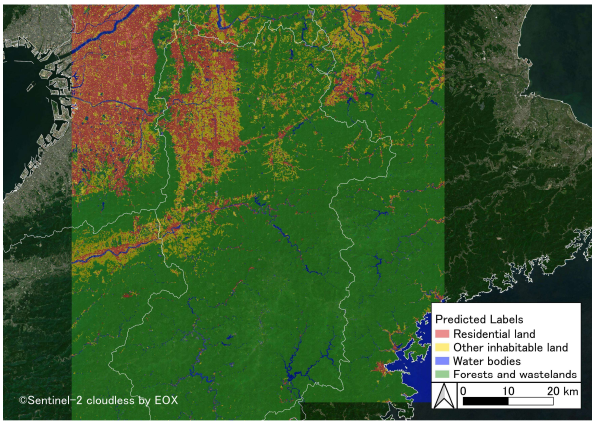

Figure 6. Land use prediction in Nara Prefecture

Nara Prefecture	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Residential land	-	99	100	99
Other inhabitable land	-	96	100	98
Water bodies	-	78	96	86
Forests and wastelands	-	100	98	99
Overall	98.3	93.2	98.4	95.6

Table 5. Evaluation metrics in Nara Prefecture

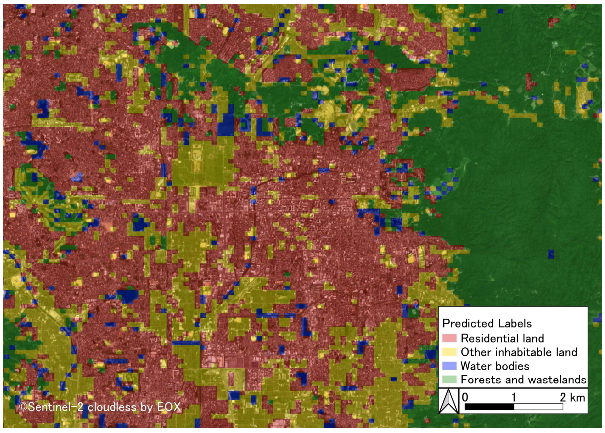

Figure 7. Predicted results in Central Nara City

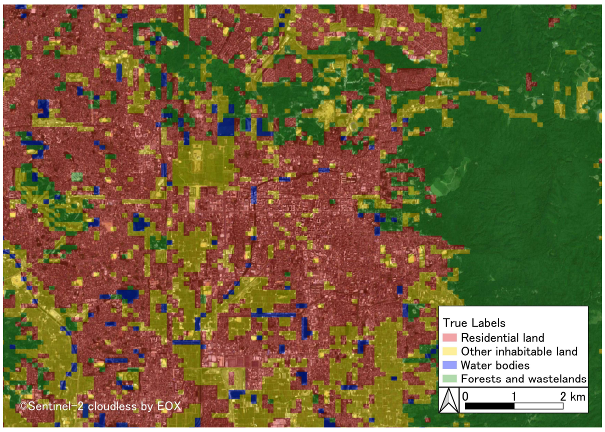

Figure 8. Ground truth labels in Central Nara City

According to Table 5, the classification performance in Nara Prefecture yielded an exceptionally high overall accuracy of 98.3%. In Nara, urban areas are concentrated in the northern part of the prefecture, while the southern part is dominated by forested regions. Compared to Akita Prefecture, this clearer distinction between inhabitable and uninhabitable areas likely contributed to the high performance. Focusing on each class, most evaluation metrics achieved high values around 90%, indicating that the model exhibits strong generalizability. On the other hand, the metrics for water bodies were slightly lower. This may be due to the presence of infrastructure overlapping rivers within some mesh units, as well as the model's tendency to overestimate water bodies around scattered ponds and lakes in urban areas.

These results indicate that the proposed model tends to misclassify areas located at the boundaries between different land use types. One possible reason is that while humans can interpret land use flexibly by considering surrounding context, machine learning models make decisions independently for each mesh unit, often relying excessively on local features. Therefore, improving classification performance in such complex boundary regions remains a challenge.

4.2 Feature Importance Analysis

To understand the behavior of the model, feature importance was analyzed for each classification stage. The visualized feature importances for all classification models are shown in Figure 9.

In the inhabitable vs. uninhabitable classification, visible light bands such as Green and Blue showed high contributions, with the Green band exhibiting the highest importance. This likely reflects its effectiveness in distinguishing urban areas and farmland based on their reflectance characteristics.

In the classification between residential land and other inhabitable land, the NIR band—which captures vegetation reflectance—had a high level of importance. On the other hand, in the classification between water bodies, and forests and wastelands, NDVI stood out with particularly high importance, demonstrating its effectiveness for this task.

Although MNDWI and NDSI did not show high importance overall, they contributed to reducing misclassifications in specific classes. In future studies, it is expected that their effectiveness can be quantitatively evaluated through more detailed classification schemes and localized accuracy assessments.

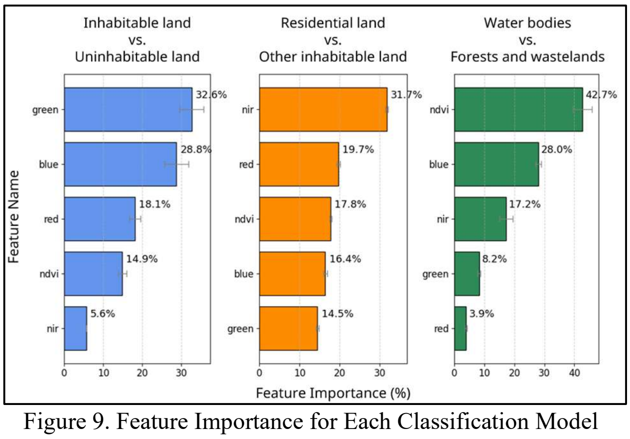

Figure 9. Feature Importance for Each Classification Model

4.3 Practical Applicability of the Proposed Method

Focusing on the processing time of the proposed method, model construction required approximately one hour, and the prediction process took about 0.5 hours each for Akita and Nara Prefectures. Considering that traditional manual interpretation methods have required years of effort and significant manpower to develop nationwide land use datasets, the introduction of this method has the potential to significantly reduce both time and costs.

Furthermore, this method achieved high classification accuracy even when using satellite imagery acquired at different time periods. This result suggests that variations in the timing of image acquisition have limited impact on classification performance, demonstrating the method's operational flexibility—an advantage in practical applications.

In addition, when combined with the high-frequency observation capabilities of Sentinel-2, this method holds the potential to enhance the timeliness of land use data updates, enabling the development of a system that can respond quickly and continuously to temporal changes. For example, even in the event of large-scale disasters that alter land use, it would be possible to rapidly identify the affected areas without the need for on-site inspections.

5. Conclusion

This study aimed to develop a novel method for automatically performing land use classification at a 100-meter mesh resolution across the entirety of Japan, a capability that had not been previously established. By leveraging satellite imagery and machine learning techniques, we successfully developed the world's first system capable of efficiently producing high-precision and high-frequency land use data. Specifically, the system was applied to Akita and Nara Prefectures to automate classification into four categories: residential land, other inhabitable land, water bodies, and forests and wastelands.

In the validation of the model constructed for Akita Prefecture, an overall accuracy of 88.2% was achieved. A closer look at class-specific performance revealed an overfitting tendency for the residential land class, indicating that the number of training samples has a significant influence on model performance. In the extrapolation test conducted for Nara Prefecture, most evaluation metrics exceeded 90%, demonstrating the model's strong generalizability across different regions and time periods.

Furthermore, this study introduced spectral indices such as NDVI, MNDWI, and NDSI as features, which contributed to improved accuracy in distinguishing water bodies and artificial surfaces. The incorporation of these indices into the feature set proved effective in enhancing the model's discrimination capabilities compared to conventional approaches.

Moreover, the proposed method requires only a short time for model training and prediction, allowing for significant efficiency gains over traditional manual processing. Given its ability to generate land use data with high frequency, this approach is expected to support various forms of decision-making in smart city contexts, including urban planning, infrastructure management, and disaster prevention. In particular, it can facilitate the rapid understanding of changes at urban and regional scales, thereby enabling more effective and practical policy development and area management.

For future work, improving classification accuracy remains a key challenge. Further efforts should include expanding beyond the four-class scheme established in this study to enable more detailed land use categorization, as well as performing classification at higher spatial resolutions. Achieving these goals will require the use of higher-resolution imagery and the application of deep learning approaches. Additionally, for areas with low classification confidence, incorporating traditional manual interpretation as a supplementary measure could enhance the overall reliability of the classification results. By validating these approaches and scaling up to cover the entire country, it will be possible to develop a more accurate and continuously updated national land use dataset.

References

Anderson, J.R., Hardy, E.E., Roach, J.T., Witmer, R.E., 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper, 964. U.S. Government Printing Office, Washington, D.C. doi.org/10.3133/pp964.

Breiman, L., 2001. Random forests. Machine Learning, 45(1), 5?32. doi.org/10.1023/A:1010933404324.

Campos-Taberner, M., García-Haro, F.J., Grau-Muedra, G., et a l., 2020. Understanding deep learning in land use classification b ased on Sentinel-2 time series. Scientific Reports, 10, 9533. doi. org/10.1038/s41598-020-74215-5.

Chen, J., Jiang, H., Zhang, W., 2021. Detection of ground mater ials using normalized difference indices with a threshold: Risk a nd ways to improve. Remote Sensing, 13(3), 450. doi.org/10.33 90/rs13030450.

ESA (European Space Agency), 2015. Sentinel-2 User Handboo k. ESA Standard Document. Available at: https://sentinel.esa.int /documents/247904/685211/sentinel-2_user_handbook.

European Commission, 2023. Sentinel-2 Global Mosaic Hub. C opernicus Land Monitoring Service. Available at: https://s2gm.l and.copernicus.eu/mosaic-hub (Accessed: April 6, 2025).

FAO (Food and Agriculture Organization), 2025. Land Cover C lassification System: Classification Concepts and User Manual. Version 3. FAO, Rome. https://www.fao.org/land-water/land/lan d-governance/land-resources-planning-toolbox/category/details/ en/c/1036361/.

Fujimoto, K., 2013: Land-use classification and the concept of i nhabitable land. Presentation at Canon Institute for Global Studi es. https://cigs.canon/event/report/uploads/pdf/130902_fujimoto _presentation.pdf.

Ghayour, L., Neshat, A., Paryani, S., Shahabi, H., Shirzadi, A., Chen, W., et al., 2021. Application of Machine Learning Approa ches for Land Cover Classification Using Sentinel-2 Data: A Co mparative Study. Remote Sensing, 13(7), 1349. doi.org/10.3390 /rs13071349.

Gong, M., Zhang, H., Yuan, Y., et al., 2020. Remote sensing im age scene classification: Benchmark and state of the art. Proceed ings of the IEEE, 109(4), 520–544. doi.org/10.1109/JPROC.202 0.3047279.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. LightGBM: A highly efficient gradient boo sting decision tree. Advances in Neural Information Processing Systems, 30, 3146?3154.

Khan, N.A., et al., 2024. A transformer-based explainable mode l for land use and land cover classification using satellite imager y. Scientific Reports, 14, 3046. doi.org/10.1038/s41598-024-671 86-4.

Koarai, M., Nakano, T., 2017. A study on the effective dissemin ation of disaster prevention and environmental information usin g geospatial data (2nd year). Technical Report of the Geospatial Information Authority of Japan, No. 19, 1–18. Available at: http s://www.gsi.go.jp/common/000205323.pdf. [in Japanese]

Li, X., Zhang, H., Sun, Y., et al., 2018. Deep learning for remote sensing image classification: A survey. WIREs Data Mining an d Knowledge Discovery, 8(6), e1264. doi.org/10.1002/widm.12 64.

MLIT (Ministry of Land, Infrastructure, Transport and Tourism). (2024). Guidelines for the Maintenance and Provision of Nation al Land Numerical Information (Draft). Tokyo, Japan. (in Japan ese). Retrieved from https://www.mlit.go.jp/tochi_fudousan_ke nsetsugyo/chirikukannjoho/content/001756426.pdf (Accessed: April 6, 2025).

Ochi, S., 2009. A study on land cover classification using object -based methods with high-resolution satellite imagery. Southeas t Asian Studies, 46(4), 578–592. [in Japanese]

Ohashi, H., Fukasawa, K., Ariga, T., Matsui, T., Hijioka, Y., 2019. High-resolution national land use scenarios under a shrinking population in Japan. Transactions in GIS, 23(2), 252?267. doi.org/10.1111/tgis.12525.

Takayanagi, S., 2017. Relationship between population dynamics and land use changes using National Land Numerical Information Land Use Detailed Mesh Data. J. City Plan. Inst. Jpn., 52(3), 459?466. doi.org/10.11361/journalcpij.52.459. [in Japanese]

Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A., Rahman, A., 2020. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. Remote Sensing, 12(7), 1135. doi.org/10.3390/rs12071135.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8(2), 127–150. doi.org/10.1016/0034-4257(79)90013-0.

Xu, H., 2006. Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. International Journal of Remote Sensing, 27(14), 3025–3033. doi.org/10.1080/01431160600589179.

Zhao, Q., Ma, T., Yang, Z., Wang, X., Yang, Y., 2019. A novel framework for urban land use classification with remote sensing and social sensing data. ISPRS Journal of Photogrammetry and Remote Sensing, 156, 247–260. doi.org/10.1016/j.isprsjprs.2019.08.018.

Zhu, X.X., Tuia, D., Mou, L., et al., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), 8–36. doi.org/10.1109/MGRS.2017.2762307.